Feature Tracking in Numerical Simulation Datasets: CMCS & Data Mining

Wendy S. Koegler

wkoegler@ca.sandia.gov

Sandia National Laboratories

June 2002

Sandia National Laboratories | Pacific Northwest National Laboratory, Operated by Battelle for the U.S. Department of Energy | Argonne National Laboratory | Los Alamos National Laboratory | NIST National Institute of Standards and Technology | MIT Massachusetts Institute of Technology | Berkeley University of California

Sharing Data & 'Information'

- Motivation: Further scientific understanding
  - Data expensive to generate
  - Identify areas that need more research
  - Compare results with more context
- Challenges:
  - Multitude of different data formats
  - Large data difficult to move
  - Various levels of search criteria (e.g. Which species?, What code?, Input summaries, Results summaries, ...)

High-Level Goals

- Enable many people to mine data created by a few.
- Enable search for data using metadata:
  - Pedigree
  - Scientific Pedigree
  - Features
- Enable data analysis & comparisons.

Definitions

- Information: High level summaries and conclusions about data
- Numerical Simulation Dataset: Usually a time-series of variables over a spatial domain.
- DNS (Direct Numerical Simulation): a simulation that uses detailed physics & chemistry
- Feature: a high level phenomena, (e.g. a vortex, flame front, crumple zone, …)
- Tracking: identification of features in time & space

Features Have Many Uses

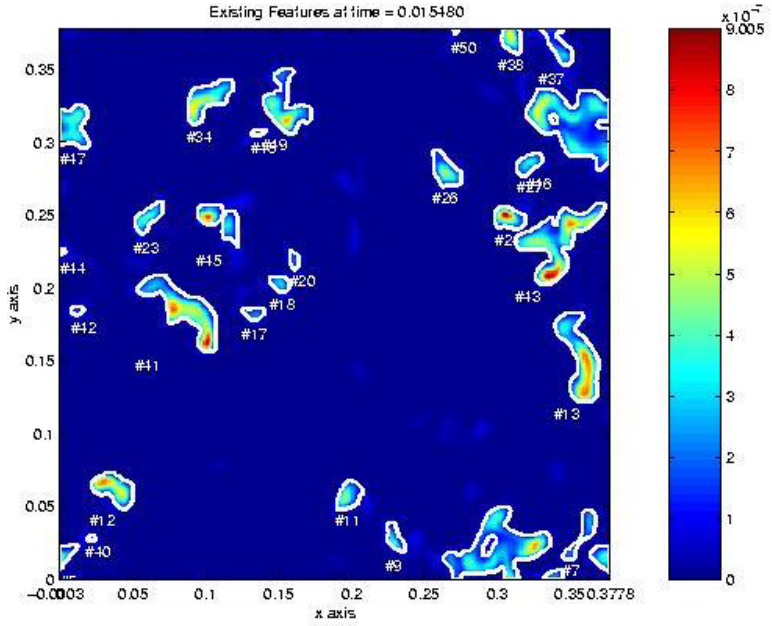
- Improved quantification of ‘Information’
- Increase automation of data analysis, data comparisons & model validation.
- Feature data becomes searchable metadata
  - Feature characteristics
  - Events
- Enhance visualization
- Reduced bandwidth requirements for data browsing & fetching

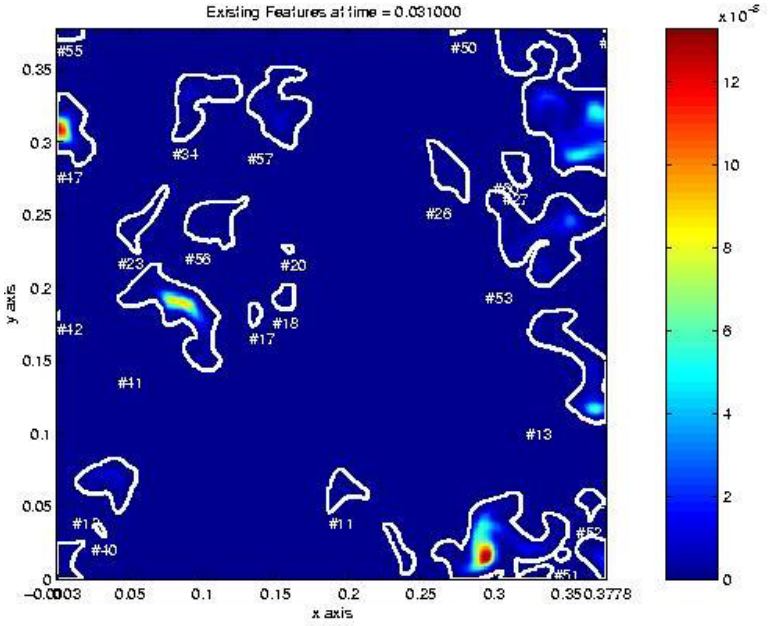
Feature Tracking Example

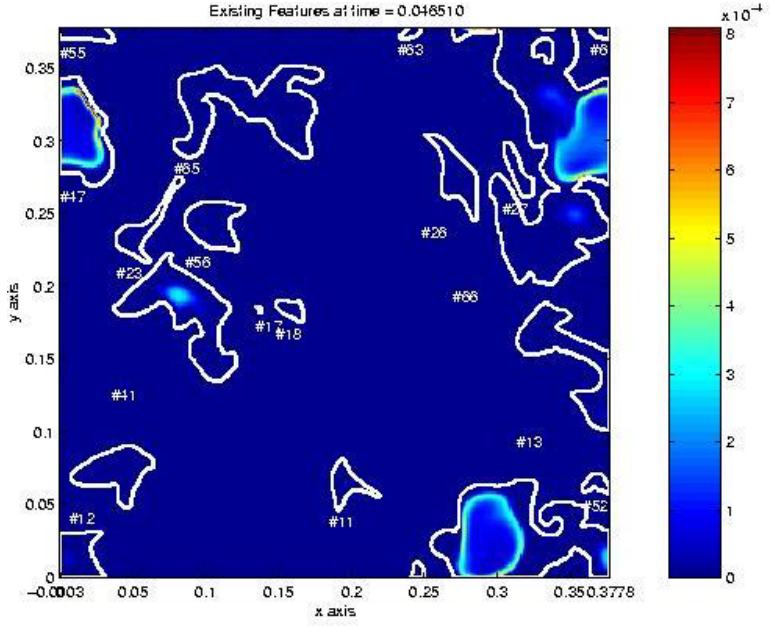
Autoignition:

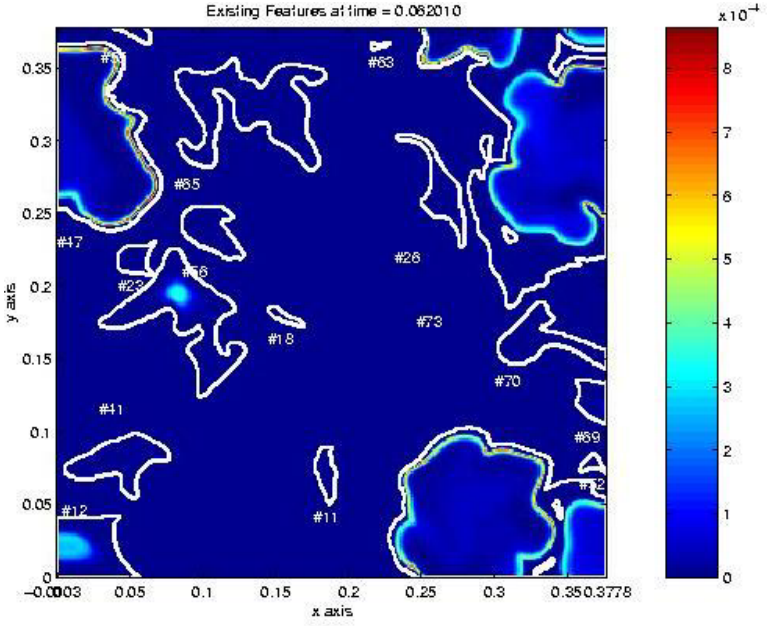
- Inhomogenous mixture
  - Hydrogen fuel (50% H_2/50% N_2) at 300K
  - Oxidizer (air) (21% O_2/79% N_2) at 1200K
- 300 x 300 grid of 0.378 x 0.378 cm (2D)
- 69 time steps with interval ~0.001 s
- Variables: Temperature and concentrations of chemical species …
- Features defined as HO_2 conc. > 1e-7

Data provided by Jackie Chen & Tarek Echekki CRF, SNL

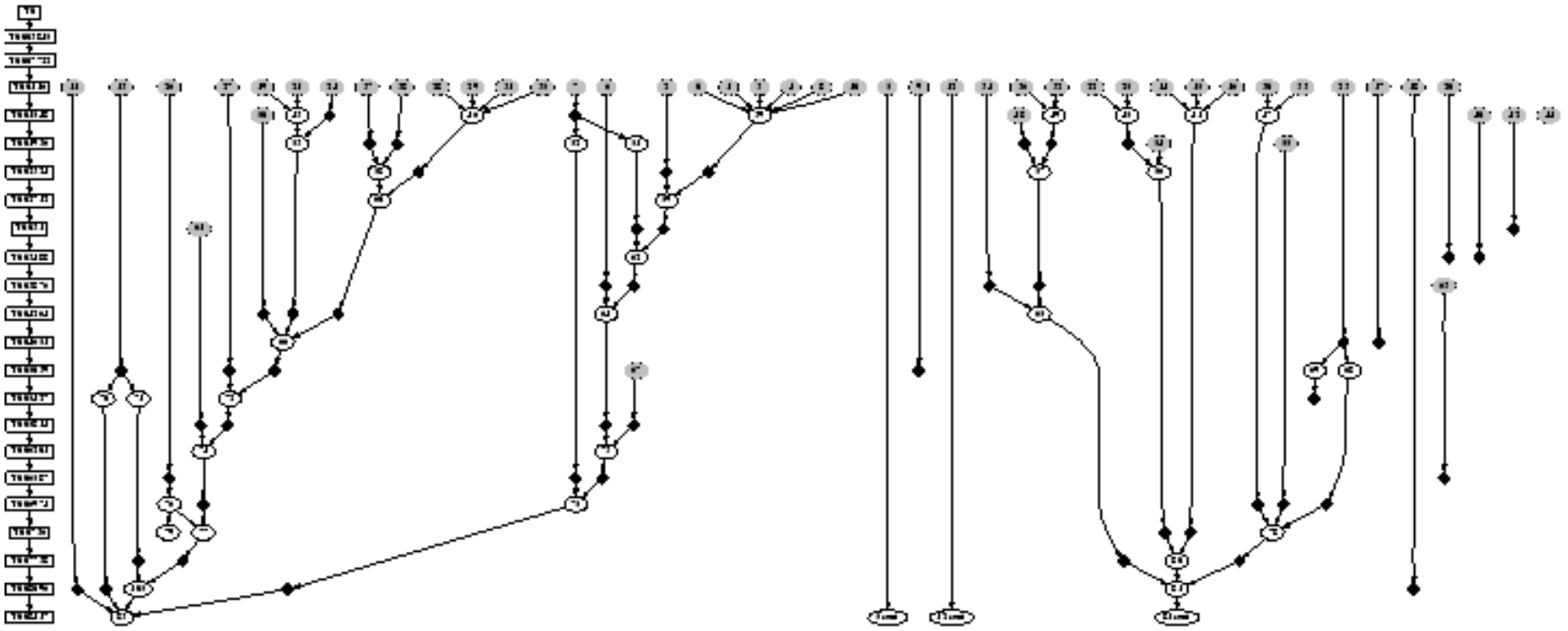

Existing Features at time = 0.015480
x 10^-4
9.005
y axis
x axis
#50
#38
#37
#34
#47
#26
#23
#15
#44
#20
#18
#12
#17
#41
#13
#11
#40
#9
#7


Existing Features at time = 0.031000
x 10^-5
y axis
x axis
#55
#50
#34
#57
#47
#26
#23
#56
#20
#53
#18
#12
#17
#41
#13
#11
#40
#52
#51


Existing Features at time = 0.046510
x 10^-4
y axis
x axis
#55
#53
#8
#65
#47
#26
#23
#56
#66
#18
#41
#13
#12
#11
#52


Existing Features at time = 0.062010
x 10^-4
y axis
x axis
#63
#65
#47
#26
#23
#56
#18
#73
#70
#41
#69
#12
#11
#2

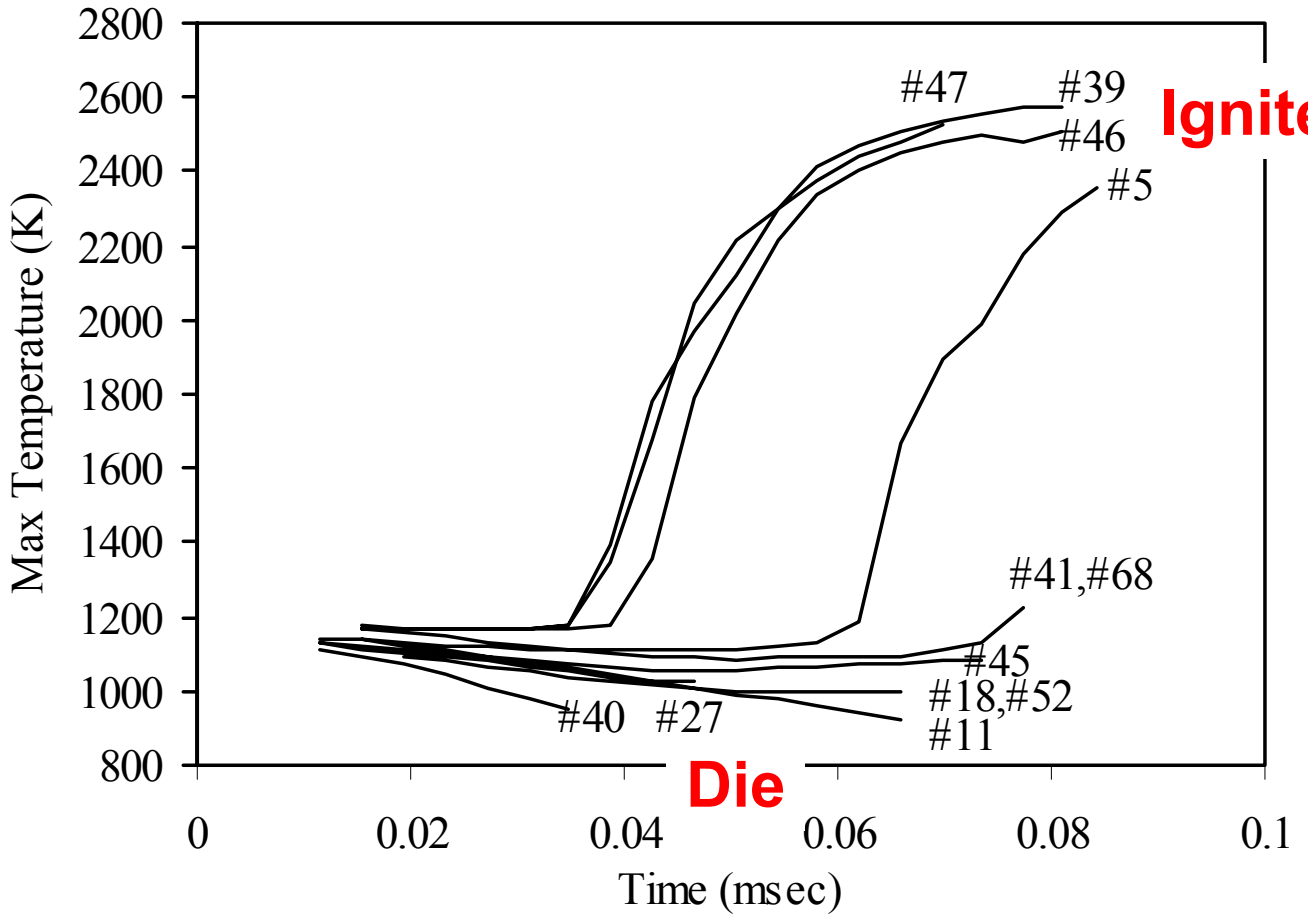
Feature Graph



Time steps

Features

Analysis: Temperature of Features



CMCS Feature Tracking Goals

- Finish development of Feature Tracking software
- Input & Reference DNS data in CMCS repository
- Annotate DNS data with feature metadata
- Configure ‘Search’ to support feature metadata
- Integrate a ‘Feature Viewer’ for feature browsing